

```

Db 121 RGDHTRAPGRHETKDOEYDSDGDMWSLMSVCSYTCGNNGKRTSCGACATATS 180
Qy 254 RTDDRNGRGEIDETRTATVEVSLAGSEFNATKLFVDTSCFEMWCKSEFLKKYH 313
Db 181 RTDRNCRIEIDETRTATVEVSLAGSEFNATKLFVDTSCFEMWCKSEFLKKYH 240
Qy 314 KYANDPSCPCSYTEVAVSTADIEDIRKDRFMKADAGPKETLEYKPTARCYRSMI 373
Db 241 KYANDPSCPCSYTEVAVSTADIEDIRKDRFMKADAGPKETLEYKPTARCYRSMI 300
Qy 374 SLESTLLAACHCCYCDNMOLITRGAGTGNLSTFESALHRYKVDYLPWITICKDMSRY 433
Db 301 SLESTLLAACHCCYCDNMOLITRGAGTGNLSTFESALHRYKVDYLPWITICKDMSRY 360
Qy 434 NEARPPNNGOKCTESPSEDYIKOFQEARX 464
Db 361 NEARPPNNGOKCTESPSEDYIKOFQEARX 391

```

```

RESULT 2
ID 095432 PRELIMINARY: PRI: 658 AA.
AC 095432
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOHETICAL 72.5 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN 11
RP SEQUENCE FROM N.A.
RA Multigenebase Sequencing Group,
RT Complete sequence of the gene for serine palmitoyltransferase,
RT subunit II found on human chromosome 14.
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RA Multigenebase Sequencing Group;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF111168; AAD09622.1;
DR InterPro: IPR000884; TSP1.
DR Pfam: PF00090; tsp.1; 1.
DR SMART: SM00209; TSP1; 1.
DR PROSITE: PS50092; TSP1; 1.
DR Hypothetical protein.
KW HYPOHETICAL protein.
SQ SEQUENCE 658 AA; 72519 MW; 7E1DF224645B300F CRC64;

```

```

Query Match 36.8%; Score 926; DB 4; Length 658;
Best Local Similarity 43.1%; Pred. No. 7.5e-81;
Matches 196; Conservative 61; Mismatches 126; Indels 72; Gaps 16;

Qy 73 DQQAHPQPRPRROETG---HPSLRDF-----PRS-----102
Db 213 DQQAHPQPRPRROETG---HPSLRDF-----PRS-----102
Qy 103 FLIDLPRPPLSKADINGQXPNIOVTEVVDGPRSEADKP--QHPEKPR-----SWSVPS 155
Db 273 FLIDLPRPPLSKADINGQXPNIOVTEVVDGPRSEADKP--QHPEKPR-----SWSVPS 155
Qy 156 PDWRAMW---QRLSLARANSQ-----DDY-----XYDSTSDSN---FLNPP 193
Db 332 L-WFLLMGDYKGEKRDAPGKGEKEDDYPSEDIEGDQEDKEDDEQALWFGT 390
Qy 194 RGNH--TARGHRTFEKNDQPEYDSTGEGDWSLMSVCSYTCGNNGKRTSCGACATAT 251
Db 391 DNMQGLAPGDWVF--KDSVSYD-YEPQKESWSPSCSGCSTGKQQRTPGQGTAT 447
Qy 252 ESRTCDPNCPCGIE--DTFRTATEVSLAGSEFNATKLFVDTSCFEMWCKSEFLK 309
Db 391 DNMQGLAPGDWVF--KDSVSYD-YEPQKESWSPSCSGCSTGKQQRTPGQGTAT 447

```

```

Db 448 ETRICDLPSCPDTEKDTLGLPSEWMLAR---NATMDHDDVDSCEKNLCKSDFLI 503
Qy 310 KYMKVNDLPSCPCSYTEVAVSTADIEDIRKDRFMKADAGPKETLEYKPTARCYI 369
Db 504 KYLSMNDLPSCPCSYTEVAVSTADIEDIRKDRFMKADAGPKETLEYKPTARCYI 563
Qy 370 RSMLESTLLAACHCCYCDNMOLITRGAGTGNLSTFESALHRYKVDYLPWITICKD 429
Db 564 RSMLESTLLAACHCCYCDNMOLITRGAGTGNLSTFESALHRYKVDYLPWITICKD 623
Qy 430 MSRYNEARPPNNGOKCTESPSEDYIKOFQEARX 464
Db 624 MSRYNEARPPNNGOKCTESPSEDYIKOFQEARX 658

```

```

RESULT 3
ID 09B0L4 PRELIMINARY: PRI: 60 AA.
AC 09B0L4
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE D107712.1 (NOVEL PROTEIN) (FRAGMENT).
GN BA149118.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN 11
RP SEQUENCE FROM N.A.
RA Smith M.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL050320; CAC36074.1;
FT NON-TER 1
FT 60
SQ SEQUENCE 60 AA; 6901 MW; 4EADCFD3EC333550 CRC64;

```

```

Query Match 12.8%; Score 321; DB 4; Length 60;
Best Local Similarity 100.0%; Pred. No. 1.1e-23;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 NNLNVGSDTSTSTPSLSKEAPREHLHQAAHQPPRRFOETGHPSLQDFPRSFLLD 106
Db 1 NNLNVGSDTSTSTPSLSKEAPREHLHQAAHQPPRRFOETGHPSLQDFPRSFLLD 60

RESULT 4
ID 0966P9 PRELIMINARY: PRI: 1121 AA.
AC 0966P9
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOHETICAL PROTEIN C36B7.5A.
GN C36B7.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN 11
RP SEQUENCE FROM N.A.
RA STRAIN-BRISTOL N2;
RC MEDLINE-99069613; PubMed-9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN 12
RP SEQUENCE FROM N.A.
RA STRAIN-BRISTOL N2;
RC Minx P.; Minx M.;
RA "The sequence of C. elegans cosmid C36B7.";

```

F:551-586/Domain: EGF homology <EGF>  
F:248,360,708,1067/Binding site: carbohydrate (Asn) (covalent) #status predicted

## Query Match

5.1%; Score 127.5; DB 2; Length 1170;

Best Local Similarity 23.3%; Pred. No. 0.08;  
Matches 41; Conservative 17; Mismatches 53; Indels 65; Gaps 8;

QY 151 WSPSPD-WPAMQSRSLIARANSQDQ--YYDSTSD-----DSNF 189  
DB 373 WPSDADDDGWSPEWSTSCATGNGIQGRGSCDNLNRCSSVQTRCHIQECDKRF 432  
QY 190 LNPRMDHAFGHRFFET-----KDPEYDSTDEGD-----222  
DB 433 -KODGWSHSPSSCSVTCGSDVITRIRLCNPPSPOMNGKPCGEARRETKACKDACP 491  
QY 223 -----WSLMSVCSVTCGNGNCKRTSC-----GYACT--ATPSRCDRPNCP 262  
DB 492 NGGWSMSPMDICSVTCGGGVCGRSLCNPNPFPQFGKDCVGDYENQCNKQDCP 547

## RESULT

7

thrombospondin 1 precursor - human

C:Species: Homo sapiens (man)

C:Date: 23-Aug-1987 #sequence-revision 03-Aug-1995 #text-change 17-Nov-2000

C:Accession: A26155; A34274; A30140; A25812; A05172; A42927

R:Lawler, J.; Hynes, R.O.

J. Cell Biol. 103, 1635-1648, 1986

A:Title: The structure of human thrombospondin, an adhesive glycoprotein with multiple

A:Reference number: A26155; MUID:87057617

A:Accession: A26155

A:Molecule type: mRNA

A:Residues: 1-1170 <LAN>

A:Cross-references: GB:X04665; NID:937137; PIDN:CAA28370.1; PID:937138

A:Note: Parts of this sequence, including the amino end of the mature protein, were dete

R:Lawler, J.; Hynes, R.O.

J. Cell Biol. 103, 1635-1648, 1986

A:Title: Characterization of the promoter region of the human thrombospondin gene. DNA S

A:Reference number: A34274; MUID:89291870

A:Accession: A34274

A:Molecule type: DNA

A:Residues: 1-166 <LAH>

A:Cross-references: GB:J04835

R:Hennesy, S.W.; Frazier, B.A.; Kim, D.D.; Deckwerth, T.L.; Baumgartel, D.M.; Rotwein,

J. Cell Biol. 108, 723-736, 1989

A:Title: Complete thrombospondin mRNA sequence includes potential regulatory sites in th

A:Reference number: A30140; MUID:89139590

A:Accession: A30140

A:Molecule type: mRNA

A:Residues: 1-83, 'A', '85-522, 'A', '524-1170 <HEM>

A:Cross-references: EMBL:X14787; NID:937464; PIDN:CAA32889.1; PID:937465

A:Note: Parts of this sequence, including the amino end of the mature protein, were dete

R:Kobayashi, S.; Eden-McClutchan, F.; Framson, P.; Bornstein, P.

Biochemistry 25, 8418-8425, 1986

A:Title: Partial amino acid sequence of human thrombospondin as determined by analysis c

A:Reference number: A25812; MUID:87157592

A:Accession: A25812

A:Molecule type: mRNA

A:Residues: 1-83, 'A', '85-357 <KOB>

A:Cross-references: GB:M25631; NID:9538353; PIDN:AAA36741.1; PID:9538354

R:Dixit, V.M.; Hennesy, S.W.; Grant, G.A.; Rotwein, P.; Frazier, W.A.

Proc. Natl. Acad. Sci. U.S.A. 83, 5449-5453, 1986

A:Reference number: A05172; MUID:86287276

A:Accession: A05172

A:Molecule type: mRNA

A:Residues: 1-83, 'A', '85-374, 'RC', <DTX>

A:Cross-references: GB:M14326; NID:9340005; PIDN:AAA61337.1; PID:9553801

A:Note: parts of this sequence, including the amino end of the mature protein, were dete

R:Sun, X.; Skorstengaard, K.; Mosher, D.F.

J. Cell Biol. 118, 693-701, 1992

A:Title: Disulfides modulate RGD-inhibitable cell adhesive activity of thrombospondin.

A:Reference number: A42927; MUID:92348511

A:Accession: A42927

A:Molecule type: protein

A:Residues: 987-1003 <SUN>

A:Note: Cys-992 is shown to have a free sulfhydryl

C:Gene: GDB:THBS1; TSP1; TSP

A:Cross-references: GDB:120438; OMIM:188060

A:Map position: 15q15-15q15

A:Introns: 23/1

A:Note: the list of introns may be incomplete

C:Complex: homotrimer, disulfide linked

C:Function:

A:Description: participates in cell migration and adhesion, and in platelet aggregati

C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;

C:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; tr

F:1-18/Domain: signal sequence #status predicted <IG>

F:19-1170/Product: thrombospondin 1 #status predicted <MAT>

F:317-375/Domain: von Willebrand factor type C repeat homology <VWC>

F:376-429/Domain: thrombospondin type 1 repeat homology <THR1>

F:434-490/Domain: thrombospondin type 1 repeat homology <THR2>

F:491-547/Domain: thrombospondin type 1 repeat homology <THR3>

F:551-586/Domain: EGF homology <EGF1>

F:650-689/Domain: EGF homology <EGF2>

F:926-928/Region: cell attachment (R-G-D) motif

F:171-232/Disulfide bonds: #status predicted

F:248,360,708,1067/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:270,274/Disulfide bonds: interchain #status predicted

F:1051/Binding site: eiytho-beta-hydroxyasparagine (asn) #status predicted

F:1051/Binding site: carbohydrate (Asn) (covalent) #status absent

## Query Match

5.0%; Score 127; DB 1; Length 1170;

Best Local Similarity 44.4%; Pred. No. 0.087;

Matches 24; Conservative 7; Mismatches 13; Indels 10; Gaps 2;

QY 219 GEGMSLMSVCSVTCGNGNCKRTSC-----GYACT--ATPSRCDRPNCP 262  
DB 494 GWMSPMDICSVTCGGGVCGRSLCNPNPFPQFGKDCVGDYENQCNKQDCP 547

## RESULT

8

procollagen N-endopeptidase (EC 3.4.24.14) I - bovine

N:Alternate names: procollagen N-proteinase

C:Species: Bos primigenius taurus (cattle)

C:Date: 15-Oct-1999 #sequence-revision 15-Oct-1999 #text-change 15-Oct-1999

C:Accession: T18517

R:Collige, A.; Nussgens, B.V.; Lapiere, C.M.

submitted to the EMBL Data Library, February 1996

A:Description: Cloning of the cDNA of the bovine procollagen I N-proteinase.

A:Reference number: Z18941

A:Accession: T18517

A:Status: preliminary; translated from GB/EMBL/DDB

A:Molecule type: mRNA

A:Residues: 1-1205 <COL>

A:Cross-references: EMBL:X96389; NID:e990769; PID:e228215; PIDN:CAA65253.1

A:Experimental source: skin

A:Gene: PC I-NP

C:Function:

A:Description: catalyzes cleavage of the propeptides of type I and II collagens prior

C:Keywords: hydrolase; metalloproteinase

## Query Match

4.9%; Score 124; DB 2; Length 1205;

Best Local Similarity 20.9%; Pred. No. 0.16;

Matches 64; Conservative 26; Mismatches 84; Indels 132; Gaps 15;

QY 192 PPRMDHAFGHRFFET-----TKDPEYDSTDEGWSLMS---VCSVTCGNGNCKRT 242  
DB 526 PPLDGTMCAPKCKHCKHCHLWLPDLKRD----GWMGMSPEGSGSRGCGVGFRT 580  
QY 243 SC-----GYACT--ATPSRCDRPNCPGIEDTFTATATEVSLAGSEFNATKLF 292

Db 581 QCDNPHANGRTCSGLADFDOLCNSODCPDALADFR----- 617  
 QY 293 DDDSCGRWMSCKSEFLKMYKMYNDLPSCPSYTEVAYSTADIFRIKRFKPMR--- 348  
 Db 618 -DEGCRW-----DLYFEHGDQHHMLHE 641  
 QY 349 -KDSGPKKLEIY--KPTARCTRSMTLESTLLAHCYGDMMQILTRG--KGAGT 402  
 Db 642 HRDA--KERCHLYCESKETG-----EYVSMKRMVHDGTRCSYKDAFSLCYVAGDCRKYGC 693  
 QY 403 PMLISTESAE-----LHYKVDVLPWITCKGDMRSYNAEPNNQCKTESPSDD 453  
 Db 694 DGVIGSSKEDCKCYGCGGNSHCKY-----VKGTFR-----SPKILG 731  
 QY 454 YIKRFQ 459  
 Db 732 YIKRFE 737  
 RESULT 9  
 T1371  
 hypothetical protein F25H8.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T1371; T24896  
 R:Gajdasty, S.  
 submitted to the EMBL Data Library, February 1996  
 A:Reference number: T1371  
 A:Accession: T21371  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2165 <W12>  
 A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3  
 A:Experimental source: clone F25H8  
 R:Gajdasty, S.  
 submitted to the EMBL Data Library, February 1996  
 A:Reference number: T1371  
 A:Accession: T21371  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2165 <W12>  
 A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3  
 A:Experimental source: clone T13H10  
 C:Genetics:  
 A:Gene: CESP:F25H8.3  
 A:Map position: 4  
 A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1; 81  
 Query Match  
 Best Local Similarity 20.0%; Pred. No. 0.33;  
 Matches 83; Conservative 39; Mismatches 129; Indels 164; Gaps 21;  
 QY 141 KDOHPNKRSWVSPDWR-----AMKORSLARANGDQDYXDSTDD----- 186  
 Db 1177 KPCHRSCPKYKL--GEMSQSVSCEDGWSRRVSCVSGNGTEVDMSLGTASDRASHQ 1234  
 QY 187 -SNFLNPP--RGWDHTA-----PGHRTPEIKQDPEYDSNDE--GD----- 222  
 Db 1235 TCMUGICPTFRNTDWSACSISGIGHERTECTIYRQSYDASFCDDTKMPTISQCHLL 1294  
 QY 223 ---W--SLWVCSVTCGNGNOKRTSC--GYACTATESHTCDRPNCGIEDTFRPAATE 274  
 Db 1295 PCTSMRPSHWSPCSVTCGSGISQTRSVSCPTGSEGTIVDEFCDRNTRPRLKKT----- 1347  
 QY 275 VSLAGEEENATKLEFVDISCE-----RMSCKS-----E 306  
 Db 1348 -----CEDTCDGPRVLOKLOADVPIIMATGPTACATGNGTQRR 1390  
 QY 307 FLKMYKMYNDLPSCPSYTEVAYSTADIFRIKRFKPMRDAAGPKKLEIYKPTAR 366  
 Db 1391 LKCRDH--VRDLDPDEVCNH-----LDREVSFRNCRRLDCS-----YKMAE 1430

QY 367 YCISMTLESTLLAHCYGDMMQILTRKGAAGTNNLISTERSAFL-----HYKV 418  
 Db 1431 W-----EECPATCGTHVQDSRNVTCSAEDGGRT---LKDVDCCVOKRPTARSNCRL 1480  
 QY 419 DVLP-----WTL-----CKGWSR-----YNEARPPNNGKQTE 447  
 Db 1481 EPCRKEGHSITWITGMSKCSASCGGWRRRRSVCTSSSCDETRKFKMDKME 1535  
 RESULT 10  
 T00355  
 hypothetical protein KIA0688 - human  
 C:Species: Homo sapiens (man)  
 C>Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
 C:Accession: T00355  
 R:ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, DNA Res. 5, 169-176, 1998  
 A:Title: Prediction of the coding sequences of unidentified human genes. X. The comp  
 A:Reference number: Z14142; MUID:98403880  
 A:Accession: T00355  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-837 <ISH>  
 A:Cross-references: EMBL:AB014588; NID:93327189; PIDN:BA031663.1; PID:93327190  
 A:Experimental source: brain  
 C:Genetics:  
 A:Gene: KIA0688  
 C:Superfamily: thrombospondin type 1 repeat homology  
 F:519-575/Domain: thrombospondin type 1 repeat homology <THR3>  
 Query Match  
 Best Local Similarity 4.8%; Score 122; DB 2; Length 837;  
 Matches 61; Conservative 21; Mismatches 98; Indels 104; Gaps 11;  
 QY 104 LLDLPNFPDLISKADINGQXPNIOVTIEVDGPDSADKQDQ-----PENKPSMSVSPDWR 159  
 Db 424 LIDKREP-----LHLPYTF---PKKDADADQCOLTFGPDSRHCPQLPPCAA 469  
 QY 160 AMWQRSLARANGDQDYXDSTSDSNFTLNPPRGMDHTAPGHRTEFTKQDPEYDST-- 217  
 Db 470 LWC-----SGHLN-----GHAMCQTKHSPWADGTPC 495  
 QY 218 -----DGEQWSLWVSVTCGNGNOKRTSC----- 244  
 Db 496 GBAQCMGRCILHNDQLODFNIPQAGWGPWGDSCRTGCGVQVSSRCQCTRPVPRNG 555  
 QY 245 GYACTA--TESRTCDRPNCP--GIEDTFRPAATEVSLAGSEFNATKLEFVDTSCERW- 300  
 Db 556 GKYCGRRTRFRSCNTEDCPTGSALTFR---EEGCAVYNNRTDLEKSFPGPMVWPVRYT 611  
 QY 301 -----MCKSEFLKMYKMYNDLPSCPS 325  
 Db 612 GYAPDQCKLTCQARALGYIVLEPRVVDGTPCS 645  
 RESULT 11  
 A42587  
 thrombospondin 2 precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Aug-1999  
 C:Accession: A42587; A39851  
 R:laherty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M. J. Biol. Chem. 267, 3274-3281, 1992  
 A:Title: Characterization of mouse thrombospondin 2 sequence and expression during ce  
 A:Reference number: A42587; MUID:92147683  
 A:Accession: A42587  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 1-1172 <LAH>  
 A:Cross-references: GB:L07803; GB:M87275; NID:9340421; PIDN:AAA53064.1; PID:9567241  
 A>Note: sequence extracted from NCBI backbone (NCBI:81502)